Progress In Electromagnetics Research, PIER 94, 83–104, 2009

A NEW CLASSIFIER FOR POLARIMETRIC SAR IMAGES

Y.-D. Zhang, L.-N. Wu, and G. Wei

School of Information Science and Engineering
Southeast University
China

Abstract—This paper proposes a hybrid classifier for polarimetric SAR images. The feature sets consist of span image, the H/A/α decomposition, and the gray-level co-occurrence matrix (GLCM) based texture features. Then, the features are reduced by principle component analysis (PCA). A 3-layer neural network (NN) is constructed, trained by resilient back-propagation (RPROP) method to fasten the training and early stop (ES) method to prevent the overfitting. The results of San Francisco and Flevoland sites compared to Wishart Maximum Likelihood and wavelet-based method demonstrate the validness of our method in terms of confusion matrix and overall accuracy. In addition, NNs with and without PCA are compared. Results show the NN with PCA is more accurate and faster.

1. INTRODUCTION

Different methods were employed for classification of polarimetric synthetic aperture radar (SAR) data during the past years, such as methods based on the maximum likelihood (ML) [1], artificial neural networks (NN) [2], support vector machines [3], fuzzy methods [4] and other approaches. Among these methods, the ML classifier can obtain good classifications. However, the results are sensitive to the complex Wishart distribution of the covariance [5, 6].

In this paper, the NN was adopted since the performance of NN classifiers is independent of the type of distribution while depends only on the training data and the discrimination power of the features [7, 8]. Classification accuracy depends mainly on the quality of features, which should be robust with maximum discrimination power and

Corresponding author: Y.-D. Zhang (zhangyudongnuaa@gmail.com).

must encompass most of the information available in the data [9]. Besides the usual target decomposition theorems [10], the gray level co-occurrence matrix (GLCM) is adopted to extract texture features [11].

During the classification stage, a 3-layer NN was proposed. To quicken the convergence of the NN weights and biases, resilient backpropagation (RPROP) training algorithm was adopted [12]. Principal component analysis (PCA) is employed to reduce the dimension of the generated feature vector [13]. Early stop (ES) method is employed to prevent the overfitting [14]. Comparison of results using complete versus reduced feature sets is presented which shows marginal difference in overall classification accuracy.

The structure of this paper is as follows: in the next section, we introduce the concept of Pauli decomposition. Section 3 presents the span image, H/A/α decomposition, and feature derived from GLCM. In Section 4, the feature reduction, structure of NN, RPROP train method, and ES method are described. Experiments in Section 5 use the NASA/JPL AIRSAR image of San Francisco site to show that our method performs better than Wishart Maximum Likelihood (WML) method. Experiments in Section 6 apply our method to crop classification on Flevoland site and prove it better than the wavelet-based method. Section 7 discusses the effect of PCA. Finally, Section 8 concludes this paper.

2. PAULI DECOMPOSITION

2.1. Basic Introduction

The features are derived from the multilook coherence matrix of the polarimetric SAR data [15]. Suppose S stands for the measured scattering matrix,

$$S = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{vh} & S_{vv} \end{bmatrix} = \begin{bmatrix} S_{hh} & S_{hv} \\ S_{hv} & S_{vv} \end{bmatrix} \tag{1}$$

Here S_{qp} represents the scattering coefficients of the targets, p the polarization of the incident field, q the polarization of the scattered field. S_{hv} equals to S_{vh} since reciprocity applies in a monostatic system configuration.

The Pauli decomposition expresses the scattering matrix S in the so-called Pauli basis, which is given by the following three 2×2 matrices:

$$S_a = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad S_b = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \quad S_c = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \tag{2}$$

Thus, S can be expressed as.

$$S = aS_a + bS_b + cS_c \tag{3}$$

where

$$a = \frac{S_{hh} + S_{vv}}{\sqrt{2}}, \quad b = \frac{S_{hh} - S_{vv}}{\sqrt{2}}, \quad c = \sqrt{2}S_{hv} \tag{4}$$

The meanings of S_a, S_b, and S_c are listed in Table 1.

The coherence matrix is obtained as:

$$T = [a, b, c][a, b, c]^T = \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{12}^* & T_{22} & T_{23} \\ T_{13}^* & T_{23}^* & T_{33} \end{bmatrix} \tag{5}$$

The average of multiple single-look coherence matrices is the multi-look coherence matrix.

2.2. Pauli Color-coded Representation

The Polarimetric information of the scattering matrix could be represented by the combination of intensities ($|S_{hh}|^2$, $|S_{vv}|^2$, $2|S_{hv}|^2$) in a single RGB image. However, the main drawback is the physical

Table 1. Pauli bases and their corresponding meanings.

Pauli Basis	Meaning
S_a	Single- or odd-bounce scattering
S_b	Double- or even-bounce scattering
S_c	Those scatterers which are able to return the orthogonal polarization to the one of the incident wave (forest canopy)

Figure 1. An example of Pauli color-coded representation.

interpretation of the resulting image in terms of $|S_{hh}|^2$, $|S_{vv}|^2$, $2|S_{hv}|^2$. Consequently, an RGB image could be formed with the intensities $|a|^2$, $|b|^2$, $|c|^2$, which correspond to clear physical scattering mechanisms as shown in Table 1.

The most employed codification corresponds to

$$|b|^2 \rightarrow \text{Red}, \quad |c|^2 \rightarrow \text{Green}, \quad |a|^2 \rightarrow \text{Blue} \tag{6}$$

Figure 1 presents an example in San Francisco area; the size of which is 1024×900.

3. FEATURE EXTRACTION

The proposed features can be divided into four types, which are explained below. The texture features consist of 4 GLCM-based features, which should be multiplied by 3 since there exist 3 channels (T_{11}, T_{22}, T_{33}) shown in Fig. 1. In addition, there are 1 span feature and $6H/\alpha$ parameters. In all, the total features are $1 + 6 + 4 \times 3 = 19$.

3.1. Span and H/A/Alpha Decomposition

The span or total scattered power is given by

$$M = |S_{hh}|^2 + |S_{vv}|^2 + 2\,|S_{hv}|^2 \tag{7}$$

Cloude and Potter [16] proposed an algorithm to identify in an unsupervised way polarimetric scattering mechanisms in the H-α plane. The method extends the two assumptions of traditional ways: 1) azimuthally symmetric targets; 2) equal minor eigenvalues λ_2 and λ_3 [17].

T can be rewritten as:

$$T = U_3 \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} U_3^H \tag{8}$$

$$U_3 = \begin{bmatrix} \cos\alpha_1 & \cos\alpha_2 & \cos\alpha_3 \\ \sin\alpha_1 \cos\beta_1 \exp(i\delta_1) & \sin\alpha_2 \cos\beta_2 \exp(i\delta_2) & \sin\alpha_3 \cos\beta_3 \exp(i\delta_3) \\ \sin\alpha_1 \sin\beta_1 \exp(i\gamma_1) & \sin\alpha_2 \sin\beta_2 \exp(i\gamma_2) & \sin\alpha_3 \sin\beta_3 \exp(i\gamma_3) \end{bmatrix} \tag{9}$$

Then, the pseudo-probabilities of the T matrix expansion elements are defined as

$$P_i = \frac{\lambda_j}{\sum_{j=1}^{3} \lambda_j} \tag{10}$$

The entropy indicates the degree of statistical disorder of the scattering phenomenon. It can be defined as:

$$H = \sum_{i=1}^{3} -P_i \log_3 P_i \quad 0 \le H \le 1 \tag{11}$$

For high entropy values, a complementary parameter (anisotropy) is necessary to fully characterize the set of probabilities. The anisotropy is defined as the relative importance of the second scattering mechanisms [18].

$$A = \frac{P_2 - P_3}{P_2 + P_3} \quad 0 \le A \le 1 \tag{12}$$

The four estimates of the angles are easily evaluated as

$$[\bar{\alpha}, \bar{\beta}, \bar{\delta}, \bar{\gamma}] = \sum_{i=1}^{3} P_i[\alpha, \beta, \delta, \gamma] \tag{13}$$

Thus, vectors from coherence matrix can be represented as (H, A, $\bar{\alpha}$, $\bar{\beta}$, $\bar{\delta}$, $\bar{\gamma}$).

3.2. Texture Features

The GLCM is a text descriptor which takes into account the specific position of a pixel relative to another. The GLCM is a matrix whose elements correspond to the relative frequency of occurrence of pairs of gray level values of pixels separated by a certain distance in a given direction [19]. Formally, the elements of a GLCM $G(i, j)$ for a displacement vector (a, b) are defined as

$$G(i, j) = |\{(x, y), (t, v) : I(r, s) = i \,\&\, I(t, v) = j\}| \tag{14}$$

where $(t, v) = (x + a, y + b)$, and $| \bullet |$ is the cardinality of a set. The displacement vector (a, b) can be rewritten as (d, θ) in polar coordinates.

GLCMs are suggested to calculate from four displacement vectors with $d = 1$ and $\theta = 0°$, $45°$, $90°$, and $135°$ respectively. In this study, the (a, b) are chosen as $(0, 1)$, $(-1, 1)$, $(-1, 0)$, and $(-1, -1)$ respectively, and the corresponding GLCMs are averaged.

The four features are extracted from normalized GLCMs; the sum of which is equal to 1. Suppose the normalized GLCM value at (i, j) is $p(i, j)$, and their detailed definition are listed in Table 2.

Table 2. Properties of GLCM.

Property	Description	Formula
Contrast	Intensity contrast between a pixel and its neighbor	$\sum_{i,j} \lvert i-j \rvert^2 p(i,j)$
Correlation	Correlation between a pixel and its neighbor (μ denotes the expected value, σ denotes the standard variance)	$\sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)\, p(i,j)}{\sigma_i \sigma_j}$
Energy	Energy of the whole image	$\sum_{i,j} p(i,j)^2$
Homogeneity	Closeness of the distribution of GLCM to the diagonal.	$\sum_{i,j} \frac{p(i,j)}{1+\lvert i-j \rvert}$

4. CLASSIFICATION METHODOLOGY

4.1. Feature Reduction

Excessive features increase computation times and storage memory. Furthermore, they sometimes make classification more complicated, which is called the curse of dimensionality. It is required to reduce the number of features [20].

Principal component analysis (PCA) is an efficient tool to reduce the dimension of a data set consisting of a large number of interrelated variables while retaining most of the variations. It is achieved by transforming the data set to a new set of ordered variables so that the first few retain most of the variations in all of the original variables. Detailed information about PCA could be seen in [21]

4.2. Neural Network Structure

Neural networks are widely used in pattern classification since they do not need any information about the probability distribution and *a priori* probabilities of different classes. A two-hidden-layer backpropagation neural network is adopted with sigmoid neurons in the hidden layers and linear neuron in the output layer.

The training vectors are formed from the selected areas and normalized and presented to the NN which is trained in batch mode [23]. The network configuration is $19 \times m \times n \times c$, i.e., a three-layer network with m neurons in the first hidden layer, n neurons in the second hidden layer, and c neuron in the output layer.

4.3. Train Method

Multilayer networks typically use sigmoid transfer functions in the hidden layers. These functions are often called "squashing" functions,

since they compress an infinite input range into a finite output range. Sigmoid functions are characterized by the fact that their slope must approach zero as the input gets large. This causes a problem when using steepest descent to train a multilayer network with sigmoid functions, since the gradient can have a very small magnitude; and therefore, cause small changes in the weights and biases, even though the weights and biases are far from their optimal values.

Riedmiller proposed RPROP training algorithm, which is to eliminate these harmful effects of the magnitudes of the partial derivatives. Only the sign of the derivative is used to determine the direction of the weight update; the magnitude of the derivative has no effect on the weight update.

The size of the weight change is determined by a separate update value. The update value for each weight and bias is increased by a factor δ_{inc} whenever the derivative of the performance function with respect to that weight has the same sign for two successive iterations. The update value is decreased by a factor δ_{dec} whenever the derivative with respect that weight changes sign from the previous iteration. If the derivative is zero, then the update value remains the same. Whenever the weights are oscillating the weight change will be reduced. If the weight continues to change in the same direction for several iterations, then the magnitude of the weight change will be increased [24].

4.4. Early Stop

One of the problems that occur during neural network training is called overfitting. The error on the training set is driven to a very small value, but when new data is presented to the network the error is large.

Early stop (ES) is a fast method that can prevent overfitting. In this technique the available data is divided into three subsets. The first one is the training subset, which is used for computing the gradient and updating the network weights/biases. The second one is the validation subset. The error on this subset is monitored during

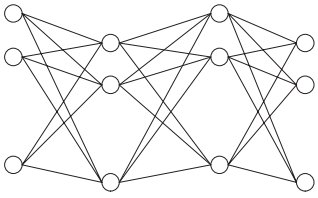


Figure 2. A three-layer neural network.

the training. The validation error normally decreases as the training set error does. However, when the network begins to overfit, the validation error typically begins to rise. When the validation error increases for a specified number of iterations, the training is stopped, and the weights/biases at the minimum of the validation error are returned. The third one is the test subset. It is used to compare different models. If the error in the test subset reaches a minimum at a significantly different iteration number compared with the validation error, it indicates a poor division of the available data.

5. TERRAIN CLASSIFICATION ON SAN FRANCISCO AREA

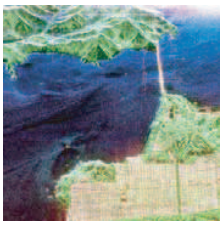
The NASA/JPL AirSAR L-band data about the San Francisco area was used for the experiments. The sub-area with size $600 * 600$ was extracted. Quantitative information about the experiment is described as follows:

- Number of features = 19
- Number of reduced features by PCA: 11 (obtained by performing PCA on total available pairs)
- Location of Sub San Francisco Area:
  - X-range: 1–600
  - Y-range: 1–600
- Parameters of GLCM
  - local area: 5×5 (pixels)
  - Number of gray levels: 8
  - Offset: [0 1]
- Properties of Training/Testing Pairs
  - Training Area
    - No. of Sea = 15810
    - No. of Urban = 9362
    - No. of Vegetated = 5064
  - Properties of Early Stop
    - TrainRatio = 0.6
    - ValidRatio = 0.2
    - TestRatio = 0.2
    - Maximum Failure Iterations = 6
  - Testing Area
    - No. of Sea = 6723
    - No. of Urban = 6800
    - No. of Vegetated = 6534
- Parameters of NN
  - Number of classes, $c = 3$
  - Configuration of neural network: $m = 10$, $n = 10$

  MB of RAM
- Hardware: Pentium 4 CPU 1.66 GHz, 512
- Software: PolSARpro v4.0, Neural Network Toolbox of Matlab 7(R14) SP3

5.1. Refine Lee Filter

The sub-area (600×600) is shown in Fig. 3(a). The refined Lee filter (Window size = 7) is used to reduce the speckle noise, and the results are shown in Fig. 3(b). The Lee filter adapts the amount of filtering to the local statistics [25]. Homogeneous areas are filtered with the maximum strength where point scatterers are let unfiltered. The refined filter could use directional windows to preserve edges and heterogeneous features [26].



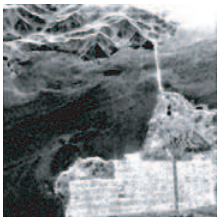
(a) The original;

(b) The refine Lee filter results

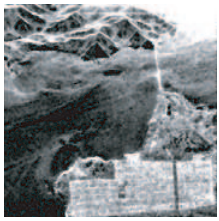
Figure 3. Pauli image of sub-area of San Francisco.

5.2. Full Features

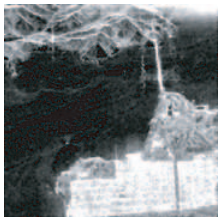
Then, the basic span image and three channels (T_{11}, T_{22}, T_{33}) are easily obtained and shown in Fig. 4. The parameters of H/A/Alpha decomposition are shown in Fig. 5. The GLCM-based parameters of T_{11}, T_{22}, T_{33} are shown in Figs. 6–8.



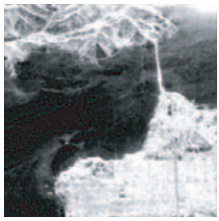
(a) Span (dB)



(b) T_{11} (dB)



(c) T_{22} (dB)



(d) T_{33} (dB)

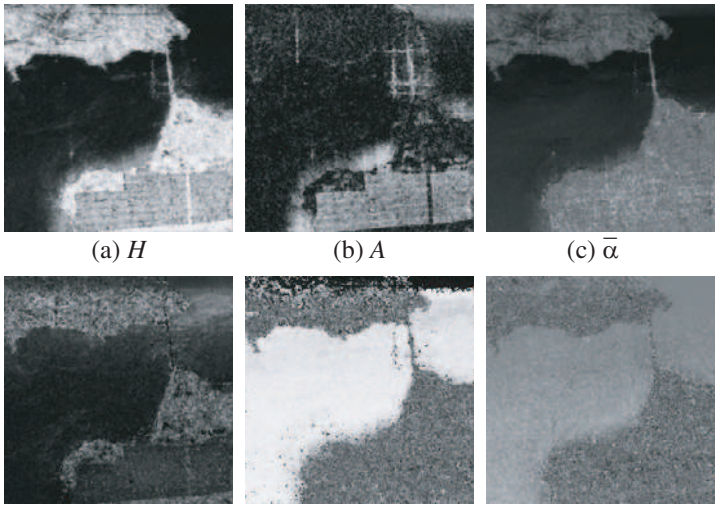
Figure 4. Basic span image and three channels image.

5.3. Feature Reduction

The curve of cumulative sum of variance with dimensions of reduced vectors via PCA is shown in Fig. 9. The detailed data are listed in Table 3. It shows that only 10 features, which are only half of the original features, could preserve 96.55% of variance.

5.4. Network Training

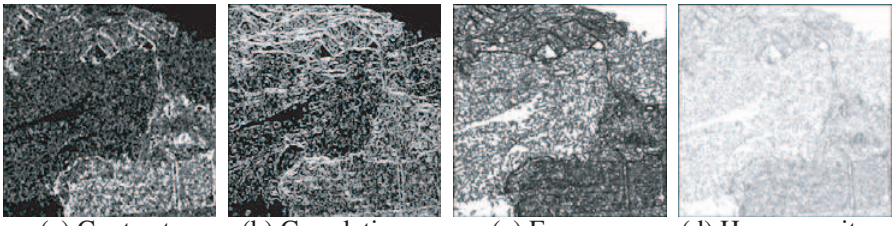
The classification is run over three classes, the sea, urban areas and vegetated zones which are selected manually shown in Fig. 10(a). The testing area is shown in Fig. 10(b).



(a) H (b) A (c) $\overline{\alpha}$

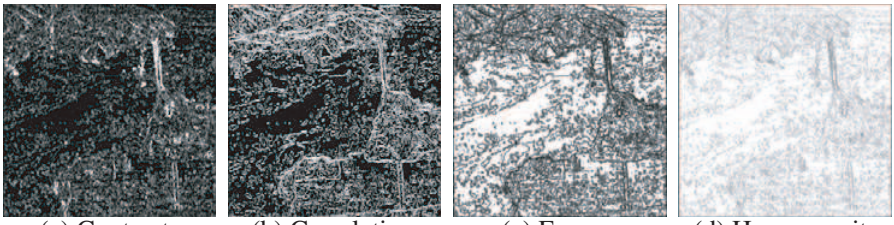
(d) $\overline{\beta}$ (e) $\overline{\delta}$ (f) $\overline{\gamma}$

Figure 5. Parameters of H/A/Alpha decomposition.



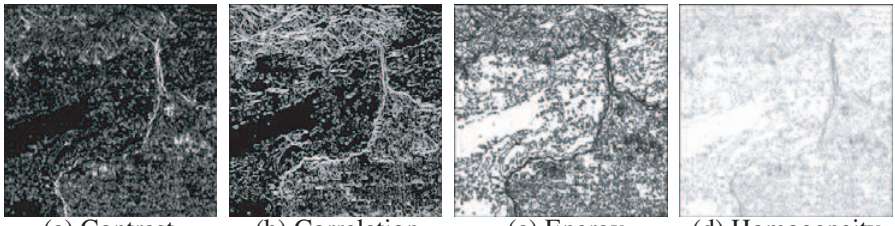
(a) Contrast (b) Correlation (c) Energy (d) Homogeneity

Figure 6. GLCM-based features of T_{11}.



(a) Contrast (b) Correlation (c) Energy (d) Homogeneity

Figure 7. GLCM-based features of T_{22}.



(a) Contrast (b) Correlation (c) Energy (d) Homogeneity

Figure 8. GLCM-based features of T_{33}.

Table 3. Detailed data of PCA on 19 features.

Dimensions	1	2	3	4	5	6	7	8	9
Variance (%)	39.78	54.40	63.17	71.30	78.94	84.54	88.49	91.99	94.59
Dimensions	**10**	11	12	13	14	15	16	17	18
Variance (%)	**96.55**	97.51	98.41	98.99	99.43	99.68	99.92	99.96	99.99

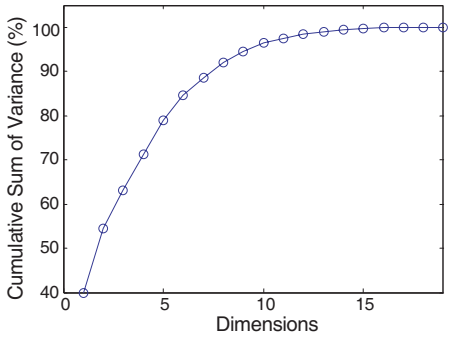
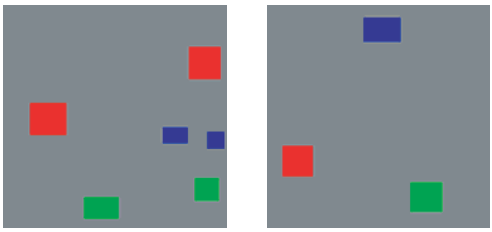


Figure 9. The curve of cumulative sum of variance with dimensions.

The training area is divided into 3 subsets: train, validation, and test subsets. Their corresponding ratios are: 0.6, 0.2, and 0.2. The performance of the network is shown in Fig. 11.

From 69th to 75th epoch, the error of train subset decreases while that of validation subset increases, which implies the over-fitting takes place, so the optimal epoch is chosen as 69. It costs 58.2734 s to train the network. Although it seems a little longer, the train process belongs to pre-processing and does not perform during classification.



(a) Training Set (b) Testing Set

Figure 10. Sample data for NN (red denotes sea, green urban areas, blue vegetated zones).

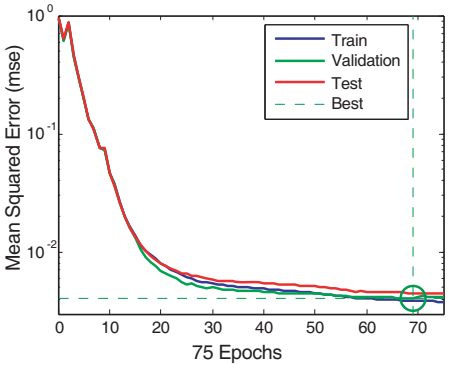


Figure 11. Performance of NN.

5.5. Classification Results

We compare our method with Wishart Maximum Likelihood (WML) statistics. The results on the whole image are shown in Fig. 12.

The confusion matrices (CM) by both methods on training and testing areas are listed in Table 4. The element of ith row and jth column represents the amount of pixels in percent belonging to class i as user defined which are assigned to class j after the supervised classification.

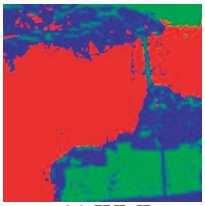
From Table 4, it is obvious that the classification accuracies in training area are all higher than 94%. For the testing area, classification accuracies are all higher than 91%. The main drawback is around vegetated zones which are easily misclassified as urban area. In all, our method is superior to the WML method because of the extended feature set and the adoption of NN.

The overall accuracies are calculated and listed in Table 5, which demonstrates that our method has a 2.2% higher overall accuracy in

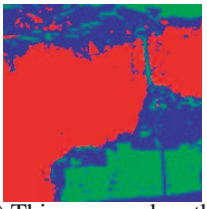
Table 4. Comparison of confusion matrix (values are given in percent).

(O denotes the output class, T denotes the target class)

		Training Area			Testing Area		
		Sea(O)	Urb(O)	Veg(O)	Sea(O)	Urb(O)	Veg(O)
WML	Sea(T)	99.95	0.01	0.04	100	0	0
	Urb(T)	0	94.17	5.83	0	97.21	2.79
	Veg(T)	0	2.43	97.57	1.07	7.65	91.27
Our Method	Sea(T)	1	0	0	99.91	0.07	0.01
	Urb(T)	0.02	99.43	0.54	0.06	98.46	1.49
	Veg(T)	0	2.41	97.59	0.90	1.58	97.52



(a) WML



(b) This proposed method

Figure 12. Classification map.

Table 5. Comparison of overall accuracy (values are given in percent).

	Training Area	Testing Area
WML	97.23	96.16
Our Method	99.42	98.64

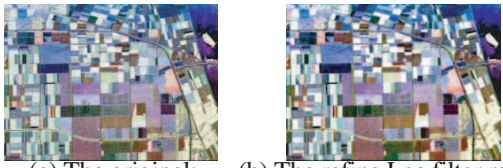
training area and 2.5% higher in testing area.

6. CROP CLASSIFICATION ON FLEVOLAND AREA

6.1. Refine Lee Filter

Flevoland, an agricultural area in The Netherlands, is chosen as another example. The site is composed of strips of rectangular agricultural fields. The scene is designated as a supersite for the earth observing system (EOS) program and is continuously surveyed by the authorities.

The Pauli image of Flevoland is shown in Fig. 13(a), and the refine Lee filtered image (Window Size = 7) is shown in Fig. 13(b).



(a) The original; (b) The refine Lee filter results

Figure 13. Pauli image of Flevoland (1024 × 750).

6.2. Full Features

Te basic span image and three channels (T_{11}, T_{22}, T_{33}) are easily obtained and shown in Fig. 14. The parameters of H/A/Alpha decomposition are shown in Fig. 15. The GLCM-based parameters of T_{11}, T_{22}, T_{33} are shown in Figs. 16–18.

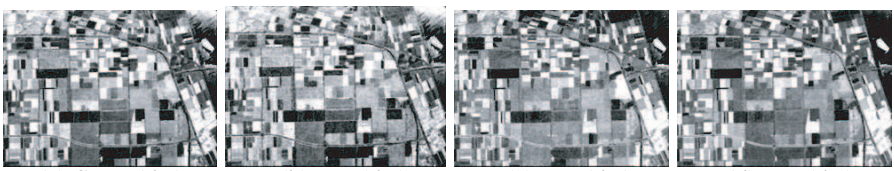
6.3. Feature Reduction

The curve of cumulative sum of variance with dimensions of reduced vectors via PCA is shown in Fig. 19. The detailed data are listed in Table 6. It shows that only 13 features, which are only half of the original features, could preserve 98.06% of variance.

6.4. Network Training

The classification is run over three classes, the sea, urban areas and vegetated zones which are selected manually shown in Fig. 20(a). The testing area is shown in Fig. 20(b).

The performance of the network is shown in Fig. 21. The optimal epoch is 254, and the training time is 31.16 s.



(a) Span (dB) (b) T_{11} (dB) (c) T_{22} (dB) (d) T_{33} (dB)

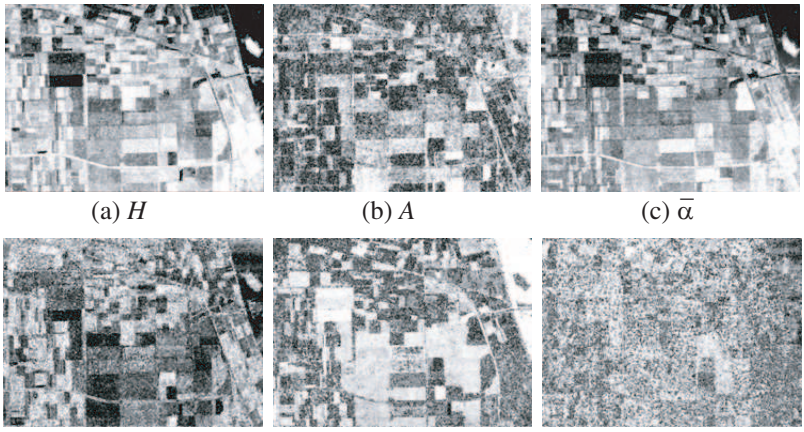
Figure 14. Basic span image and three channels image.

6.5. Classification Results

We apply our method on the whole image. The results on the whole image are shown in Fig. 22.

From Fig. 22, it is clear that our method can classify most of the areas correctly. The confusion matrices on training and testing areas are calculated and listed in Table 7.

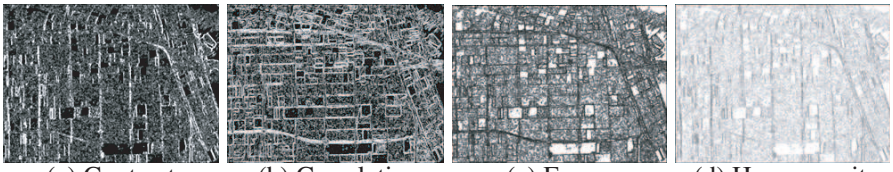
The overall accuracies of our method on train and test areas



(a) H (b) A (c) $\overline{\alpha}$

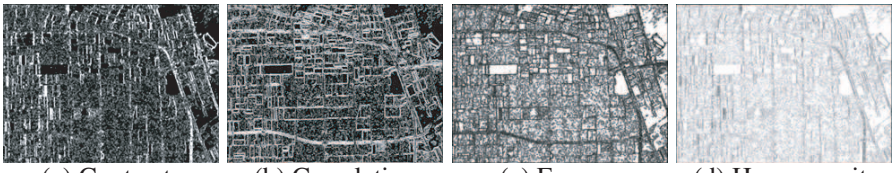
(d) $\overline{\beta}$ (e) $\overline{\delta}$ (f) $\overline{\gamma}$

Figure 15. Parameters of H/A/Alpha decomposition.



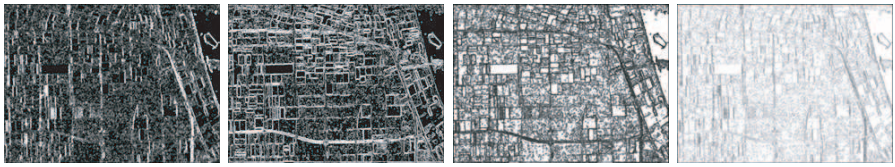
(a) Contrast (b) Correlation (c) Energy (d) Homogeneity

Figure 16. GLCM-based features of T_{11}.



(a) Contrast (b) Correlation (c) Energy (d) Homogeneity

Figure 17. GLCM-based features of T_{22}.



(a) Contrast (b) Correlation (c) Energy (d) Homogeneity

Figure 18. GLCM-based features of T_{33}.

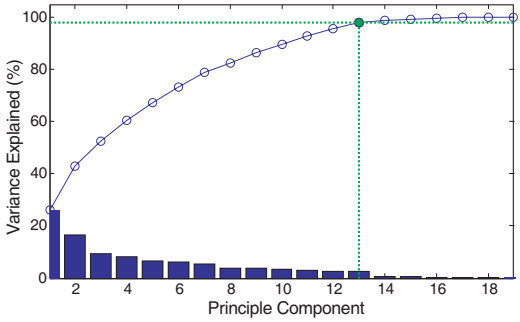
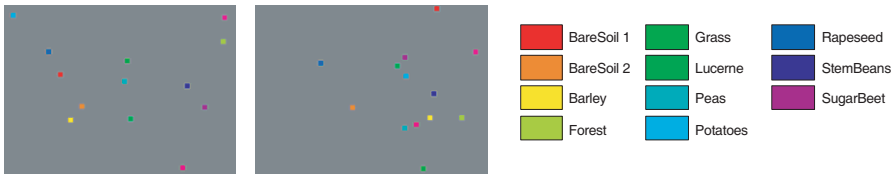


Figure 19. The curve of cumulative sum of variance with dimensions.

Table 6. Detailed data of PCA on 19 features.

Dimensions	1	2	3	4	5	6	7	8	9
Variance (%)	26.31	42.98	52.38	60.50	67.28	73.27	78.74	82.61	86.25
Dimensions	10	11	12	**13**	14	15	16	17	18
Variance (%)	89.52	92.72	95.50	**98.06**	98.79	99.24	99.63	99.94	99.97



(a) Training Set (b) Testing Set (c) Legend of Colors

Figure 20. Sample data areas for NN.

are 98.62% and 92.87% respectively. Fukuda used a wavelet-based method to classify the same Flevoland area, and it obtained the overall accuracy as 88.28% (see Table 2 in Ref. [27]).

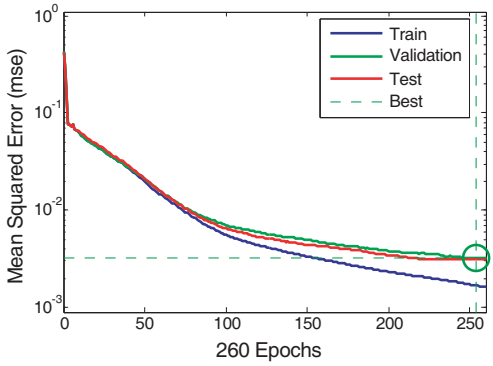


Figure 21. Performance of NN.

Table 7. Confusion matrix comparison on train area (values are given in percent).

The overall accuracy is 98.62%

Class	1(O)	2(O)	3(O)	4(O)	5(O)	6(O)	7(O)	8(O)	9(O)	10(O)	11(O)	12(O)	13(O)
1(T)	99.75	0	0	0	0	0	0	0	0	0	0	0.25	0
2(T)	0	99.75	0	0	0	0	0	0	0	0	0.25	0	0
3(T)	0	0	100	0	0	0	0	0	0	0	0	0	0
4(T)	0	0	0	93	0	0	6.25	0.5	0	0.25	0	0	0
5(T)	0	0	0.25	0	99.25	0.5	0	0	0	0	0	0	0
6(T)	0	0	0	0	0	100	0	0	0	0	0	0	0
7(T)	0	0	0	1	0	0	97.5	0.75	0	0	0.75	0	0
8(T)	0	0.25	0	3	0	0	0	96.5	0	0.25	0	0	0
9(T)	0	0	0	0	0	0	0	0	99.75	0	0	0.25	0
10(T)	0	0	0	0.25	0	0	0.25	0.25	0	99.25	0	0	0
11(T)	0	0.25	0	0	0	0	1.5	0	0	0	98.25	0	0
12(T)	0	0	0	0	0	0	0	0	0.75	0	0	99.25	0
13(T)	0	0	0	0	0	0.25	0	0	0	0	0	0	99.75

7. DISCUSSIONS ON WITH/WITHOUT PCA

7.1. Overall Accuracy

Sixty square areas of different size are picked out randomly from the Flevoland area and classified with neural networks with/without PCA. The overall accuracies of these two NNs are observed and shown in Fig. 23. It should be noted that input data of the NN without PCA should be normalized although the PCA is omitted. Otherwise the performance of NN will decrease rapidly.

Table 8. Confusion matrix comparison on test area (values are given in percent).

The overall accuracy is 92.87%

Class	1(O)	2(O)	3(O)	4(O)	5(O)	6(O)	7(O)	8(O)	9(O)	10(O)	11(O)	12(O)	13(O)
1(T)	96.5	0	0	0	0	0	0	0	0	0	3.5	0	0
2(T)	0	98.75	0	0	0	0	0.25	0	1	0	0	0	0
3(T)	3	0	94.75	0	2.25	0	0	0	0	0	0	0	0
4(T)	0	0	0	94.75	0	0	5	0	0	0.25	0	0	0
5(T)	0	0	0	0	95.5	4.5	0	0	0	0	0	0	0
6(T)	0	0	0	0	16	83.75	0	0	0	0.25	0	0	0
7(T)	0	0.25	0	15	0	0	77	6.25	0	0.25	1.25	0	0
8(T)	0	0.75	0	0	0	0	0	99.25	0	0	0	0	0
9(T)	0	0	0	0	0	0	0	0	91.25	0	7.5	0	1.25
10(T)	0	0	0	0	0	0	0	0	0	98.75	0	0	1.25
11(T)	0	0.5	0	0.5	0	2	2.75	0	0.5	0	91.75	0	2
12(T)	6.25	0	1	0	0	0	0	0	2	0	0	90.75	0
13(T)	0	2.5	0	0	0	0	0	0	0	0	3	0	94.5

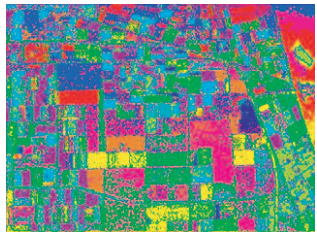


Figure 22. Classification map of our method.

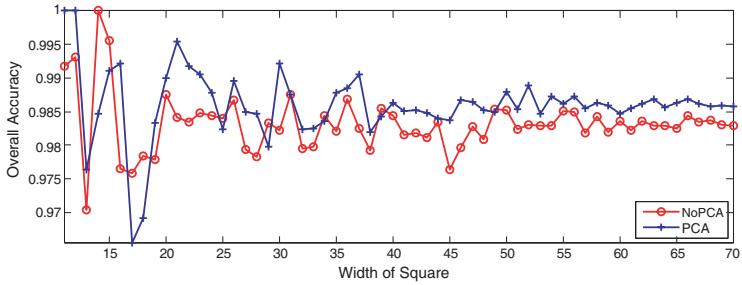


Figure 23. The overall accuracy versus square width.

The mean of overall accuracy of NN with PCA is 0.9861, and the mean of overall accuracy of NN without PCA is 0.9832. It

demonstrates that the overall accuracy of the NN with PCA is 0.003 superior to that of the NN without PCA.

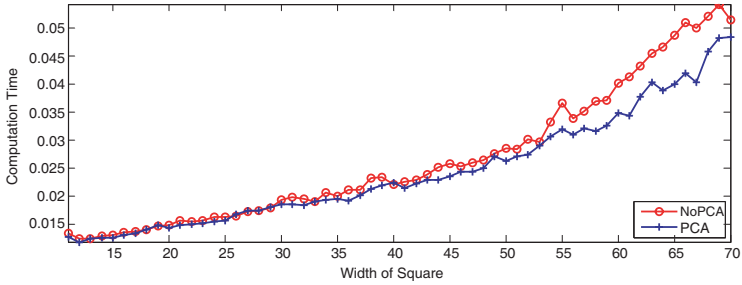


Figure 24. Computation time versus square width.

7.2. Computation Time

Computation times are depicted in Fig. 24, which indicates that NN with PCA enjoys a less computation time than NN without PCA. The time difference is gradually becoming large as the width of the randomly selected area increases.

Suppose W denotes the width of square, n the corresponding pixels, and $n = W^2$. We can calculate the time complexity of our method from Fig. 24.

$$Time = 9.118 \times 10^{-6} w^2 - 1.943 \times 10^{-4} w + 1.455 \times 10^{-2} \qquad (15)$$

The fitting curve is shown in Fig. 25. The curve fits the data quite accurately, and goodness-of-fit statistics is listed in Table 9.

It is obvious that the time complexity of our method is $O(W^2)$. Since $n = W^2$, we can conclude that $Time = O(n)$.

Table 9. Goodness of fit statistics.

Index	Value
Sum of squares due to error	9.0753×10^{-5}
Coefficient of determination	0.9843
Degrees of freedom	57
Degree-of-freedom adjusted coefficient of determination	0.9838
Root mean squared error	0.0013

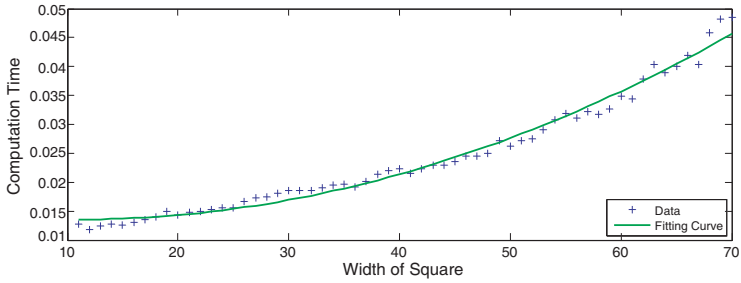


Figure 25. Time complexity of our method.

8. CONCLUSION

In this paper, a hybrid feature set has been introduced which is made up of the span image, H/A/α decomposition, and GLCM-based texture features. Then, a two-hidden-layer neural network has been trained by RPROP and ES method. The overall accuracies of the classification of San Francisco and Flevoland sites demonstrate the validity of our method. In addition, using a reduced feature set with PCA to replace the total feature set is also proved effective and faster.

REFERENCES

1. Lee, J. S., M. R. Grunes, and R. Kwok, "Classification of multi-look polarimetric SAR imagery based on complex Wishart distribution," *Int. J. Remote Sensing*, Vol. 15, No. 11, 2299–2311, 1994.
2. Yang, S. Y., M. Wang, and L. C. Jiao, "Radar target recognition using contourlet packet transform and neural network approach," *Signal Processing*, Vol. 89, No. 4, 394–409, 2009.
3. Chen, C. H. and P.-G. P. Ho, "Statistical pattern recognition in remote sensing," *Pattern Recognition*, Vol. 41, No. 9, 2731–2741, 2008.
4. Kasturirangan, K., "Space technology for humanity: A profile for the coming 50 years," *Space Policy*, Vol. 23, No. 3, 159–166, 2007.
5. Makal, S., A. Kizilay, and L. Durak, "On the target classification through wavelet-compressed scattered ultrawide-band electric field data and ROC analysis," *Progress In Electromagnetics Research*, PIER 82, 419–431, 2008.
6. Rostami, A. and A. Yazdanpanah-Goharriz, "A new method for classification and identification of complex fiber bragg

grating using the genetic algorithm," *Progress In Electromagnetics Research*, PIER 75, 329–356, 2007.

7. Khan, K. U. and J. Yang, "Polarimetric synthetic aperture radar image classification by a hybrid method," *Tsinghua Science and Technology*, Vol. 12, No. 1, 97–104, 2007.
8. Castaldi, G., V. Galdi, and G. Gerini, "Evaluation of a neural-network-based adaptive beamforming scheme with magnitude-only constraints," *Progress In Electromagnetics Research B*, Vol. 11, 1–14, 2009.
9. Li, X. and J. Gao, "Pad modeling by using artificial neural network," *Progress In Electromagnetics Research*, PIER 74, 167–180, 2007.
10. Cloude, S. R. and E. Pottier, "A review of target decomposition theorems in radar polarimetry," *IEEE Trans. Geosci. Remote Sensing*, Vol. 34, No. 2, 498–518, 1996.
11. Cooper, G. R. J. and D. R. Cowan, "The use of textural analysis to locate features in geophysical data," *Computers & Geosciences*, Vol. 31, No. 7, 882–890, 2005.
12. Mandal, S., P. V. Sivaprasad, S. Venugopal, and K. P. N. Murthy, "Artificial neural network modeling to evaluate and predict the deformation behavior of stainless steel type AISI 304L during hot torsion," *Applied Soft Computing*, Vol. 9, No. 1, 237–244, 2009.
13. Luukka, P., "Classification based on fuzzy robust PCA algorithms and similarity classifier," *Expert Systems with Applications*, Vol. 36, No. 4, 7463–7468, 2009.
14. Durrington, P. N., V. Charlton-Menys, C. J. Packard, et al., "Familial hypobetalipoproteinemia due to a novel early stop mutation," *Journal of Clinical Lipidology*, Vol. 2, No. 5, 384–390, 2008.
15. Shyu, J.-J., C.-H. Chan, M.-W. Hsiung, P.-N. Yang, H.-W. Chen, and W.-C. Kuo, "Diagnosis of articular cartilage damage by polarization sensitive optical coherence tomography and the extracted optical properties," *Progress In Electromagnetics Research*, PIER 91, 365–376, 2009.
16. Cloude, S. R. and E. Pottier, "An entropy based classification scheme for land applications of polarimetric SAR," *IEEE Trans. Geosci. Remote Sensing*, Vol. 35, No. 1, 549–557, 1997.
17. Duan, Y., S.-J. Lai, and T. Huang, "Coupling projection domain decomposition method and meshless collocation method using radial basis functions in electromagnetics," *Progress In Electromagnetics Research Letters*, Vol. 5, 1–12, 2008.

18. Pottier, E. and S. R. Cloude, "Application of the H/A/α polarimetric decomposition theorems for land classification," *Proc. SPIE Conference on Wideband Interferometric Sensing and Imaging Polarimetry*, 132–143, San Diego, CA, USA, 1997.
19. Tien, C. L., Y. R. Lyu, and S. S. Jyu, "Surface flatness of optical thin films evaluated by gray level co-occurrence matrix and entropy," *Applied Surface Science*, Vol. 254, 4762–4767, 2008.
20. Bermani, E., S. Caorsi, and M. Raffetto, "An inverse scattering approach based on a neural network technique for the detection of dielectric cylinders buried in a lossy half-space," *Progress In Electromagnetics Research*, PIER 26, 67–87, 2000.
21. Luukka, P., "Classification based on fuzzy robust PCA algorithms and similarity classifier," *Expert Systems with Applications*, Vol. 36, No. 4, 7463–7468, 2009.
22. Zainud-Deen, S. H., H. A. El-Azem Malhat, K. H. Awadalla, and E. S. El-Hadad, "Direction of arrival and state of polarization estimation using radial basis function neural network (RBFNN)," *Progress In Electromagnetics Research B*, Vol. 2, 137–150, 2008.
23. Zhang, Y. D. and L. Wu, "Weights optimization of neural network via improved BCO approach," *Progress In Electromagnetics Research*, PIER 83, 185–198, 2008.
24. Riedmiller, M. and H. Braun, "A direct adaptive method for faster backpropagation learning: The RPROP algorithm," *Proceedings of the IEEE International Conference on Neural Networks*, 586–591, San Francisco, 1993.
25. Lee, Y. H. and S. Y. Huang, "Electromagnetic susceptibility of an electromagnetic band-gap filter structure," *Progress In Electromagnetics Research B*, Vol. 15, 31–56, 2009.
26. Gupta, K. K. and R. Gupta, "Despeckle and geographical feature extraction in SAR images by wavelet transform," *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 62, No. 6, 473–484, 2007.
27. Fukuda, S. and H. Hirosawa, "A wavelet-based texture feature set applied to classification of multifrequency polarimetric SAR images," *IEEE Trans. on Geoscience and Remote Sensing*, Vol. 37, No. 5, 2282–2286, 1999.